DOCUMENT RESUME

ED 281 180 CS 008 791

AUTHOR Anderson, Richard C.; Davison, Alice
TITLE Conceptual and Empirical Bases of Readability Formulas. Technical Report No. 392.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Oct 86
CONTRACT 400-81-0030
NOTE 64p.
PUB TYPE Information Analyses (070)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Basal Reading; Readability; *Readability Formulas; Reader Response; *Reader Text Relationship; *Reading Comprehension; Reading Instruction; Reading Interests; *Reading Material Selection; Reading Processes; Reading Research; Research Problems

ABSTRACT

The problems arising from treating word and sentence complexity as the direct causes of difficulty in comprehension are surveyed in this paper from the perspective of readability formulas. The basic choices and assumptions made in the development and use of readability formulas are discussed in relation to the larger question of text comprehensibility. When considering how close these formulas come to being accurate and informative predictors of comprehension, it is argued that readability formulas are not the most appropriate measure and that they cannot reliably predict how well individual readers will comprehend particular texts. The paper shows that the high correlations between formula predictions based on text features reported in most readability research are the by-product of using an inappropriate statistical model that aggregates texts and readers, giving an exaggerated impression of the contribution of linguistic factors in the text concerning comprehension. Text and reader properties that cannot be measured by formulas are emphasized as having a far greater influence on comprehension. Further, it is argued that no readability formula can be a reliable guide for editing a text to reduce its difficulty. (Ten pages of references are included.) (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED281180

CENTER FOR THE STUDY OF READING

Technical Report No. 392

CONCEPTUAL AND EMPIRICAL BASES OF READABILITY FORMULAS

Richard C. Anderson

Alice Davison

University of Illinois at Urbana-Champaign

October 1986

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The work upon which this publication was based was performed pursuant to Contract No. 400-81-0030 of the National Institute of Education. It does not, however, necessarily reflect the views of the agency.

CS008791

Abstract

The question of what features make a text difficult or easy for a reader is examined in this paper, which looks at the implications from the perspective of readability formulas. This is related to the larger question of text comprehensibility. Problems arise when difficult words and long sentences are treated as the direct cause of difficulty in comprehension and are used in readability formulas to predict the readers' comprehension. Readability formulas are not the most appropriate measure and cannot reliably predict how well individual readers will comprehend particular texts. Far more important are text and reader properties which formulas cannot measure. Neither can any formula be a reliable guide for editing a text to reduce its difficulty.

Conceptual and Empirical Bases of Readability Formulas

The question of what features of a text make it easy for a reader is interesting from many different perspectives. In this paper we will examine this question and its implications from the specific perspective of readability formulas, pointing out the basic choices and assumptions made in their development and use. These assumptions will be discussed in relation to the larger question of text comprehensibility in which the use of formulas is embedded. We question to what degree readability formulas actually do what they were intended to do: to gauge whether particular texts can be read and understood by particular readers or groups of readers, on some particular use or occasion of reading.

We will argue that readability formulas are not the most appropriate measures for this purpose, for the reasons which follow. Summarizing the arguments, we note that the aggregate statistical model which readability formulas are based on is inappropriate. As a consequence, formulas do not reliably predict comprehension for individual readers. Formulas are also misleading guides for editing a text to reduce its difficulty. They measure features of a text which are at best correlated with difficulty, without being a more specific causal model. A causal model would define what features of language actually contribute directly to difficulty in comprehension, whereas formulas, being based only on statistical correlations, cannot be used to diagnose what is difficult about the language in a text.

Formulas are applied by calculating the average sentence length and word difficulty in short samples of texts. Features of a text not among the features of sentence and word difficulty almost certainly make a much greater difference to comprehension than the features which are measured by applying a formula. The criteria of comprehension associated with formulas are comprehension measures which are generally the least sensitive to specific features of language, of the experimental measures currently in use. Finally, to the extent that formulas do capture some plausible intuitions about the working memory capacity of a reader, this notion needs to be made more explicit in the context of basic research using on-line measures of attention and comprehension.

We will start by describing one of the earliest readability formulas, proposed in Vogel and Washburne (1928) and noting the characteristics which have persisted in the more modern formulas now in use. Vogel and Washburne based their study on a sample of 700 books which had been mentioned by 37,000 children as ones they had liked. The scores of these children on the paragraph meaning section of the Stanford Achievement test allowed them to be placed in grade-level rankings. The linguistic features of the books were measured and correlated with the reading scores of the children who had read and liked the books. From this information, a formula was designed which is used to predict what reading scores are necessary for a reader to read a certain book.

The Vogel and Washburne Formula consists of the following:

1) number of different words in a 1000 word sample;
2) total number of prepositions in the 1000 word sample;
3) total number of words not on the Thorndike list of the 10,000 most frequent words;
4) the number of clauses in 75 sample sentences

These factors enter into a regression equation:

Reading test score: $= .085x_1 + .101x_2 + .604x_3 - .411x_4 + 17.43$

The reading score levels which the formula predicted for books correlated .85 with the average reading test scores of the children in the sample who had read and liked the books (Chall, 1958, p. 19 and passim, Klare, 1963, p. 39).

This early formula illustrates the features which are still typical of readability formulas as a class, and it should be noted that these features represented advances in research and research methods of that period. Thorndike's (1921) list of word frequencies was the first large-scale study of English vocabulary use on an objective empirical basis. Regression equations were a new statistical procedure which allowed large amounts of data to be integrated. Standard achievement tests, which had been recently developed, provided an objective way of comparing students and ranking them. The measures of language in a text sample focused on fairly easily defined units (words, sentences, prepositions) which occur in large numbers in a text. The sample of students and books which were studied included a wide range of

variation, and the correlations of features of text and student scores were very high. Note that unlike much subsequent readability research, the books sampled were not school texts edited to a certain grade, nor short passages contrived to test reading achievement.

The early formulas, like the Vogel and Washburne formula just described, represented a considerable advance in research at that time. The concepts of formulas has undergone considerable development since 1928, but the general idea has remained the same. Some specific features have changed, however, such as methods of sampling texts and measuring comprehension. The independent measure of student performance has typically been the ability to answer correctly 50% or more of multiple choice comprehension questions, or to retrieve 30% or more of the deleted words in a cloze test. Different formulas have used different text variables and ways of counting them, but all formulas use some measure of word difficulty and of sentence complexity. (For more complete discussion of specific formulas, see the overviews in Chall (1958), Klare (1963, 1974, 1975, 1984), and the discussion of many text variables and cloze as a comprehension measure in Bormuth (1966)). The basic formulas have not changed in any fundamental way, either in the assumptions behind them, or in the way that the problem of text difficulty is conceived.

Anyone who reads surveys of formulas and the problems of measuring text difficulty will be struck by the fact that scholars who do research on readability formulas are aware of the

range of features that make a text complex or easy for a reader. These scholars present lucid and perceptive discussions of those aspects of texts and readers which are not measured by formulas, such as writing style, text organization and background knowledge of the reader (Gray & Leary, 1935; Chall, 1958, 1984; Klare, 1963, 1984, for example). These writers are quite clear about what formulas are sensitive to and what results can be expected from them. Both Chall (1958:97ff) and Klare (1963:20, 122ff.) note that efforts to increase the readability of texts by simplifying the vocabulary and sentences do not consistently lead to improved comprehension as measured by ability to answer questions, to recall important features of content, and to retain information over time. Nevertheless, both Chall and Klare interpret available evidence as demonstrating that vocabulary and sentence complexity account for a large proportion of the variance in the understanding of texts (cf. Chall, 1984, as well as Chall, 1958, Klare, 1963).

Scholars of readability are also aware of the impossibility of reducing all text or reader properties to formula variables. To accommodate formulas to the great variety in texts, they attach external conditions to formulas. These take the form of injunctions not to use the formulas for revising texts, or for assessing certain kinds of text (poetry, mathematics, unusual texts of various kinds) and not to take formula values as anything but rough predictions of text ease or difficulty. But these injunctions are not built into formulas, as an intrinsic and unavoidable part of them. It is easy to overlook hedges and

restrictions added onto a mathematical formula which has the immense lure of statistical correlation behind it.

The world at large, including publishers and purchasers of textbooks, has not heeded the responsible and well-founded warnings of writers like Chall and Klare. The formula variables--word difficulty and sentence length/complexity--look like factors that could strain a reader's capacity to process linguistic information. Writers and editors who ignore the difference between correlation and causation persist in seeing a formula as a model of what causes a text to be difficult, so that when under pressure to revise a text which might be difficult for a variety of reasons, they simplify hard words and split up complex sentences in the hope that these factors have enough causal power to make a difference in comprehension (cf. Davison & Kantor, 1982, and Green & Olsen, to be published.)

The damage done to text cannot be blamed on scholars like Chall and Klare, or even entirely on people who misunderstand the meaning of correlation. The problem is that there are no clear or widely accepted alternatives to the formula-like approach to the problem of linguistic variables and text comprehensibility, although field-testing on a sample of readers and the judgment of experienced readers are possibilities (Klare, 1984). The research on linguistic and other properties of texts which influence comprehension has not yet provided any comprehensive model of how the language of a text is understood, which would be more insightful and effective than formulas. There is, however, a substantial body of research which has made considerable

progress in illuminating important aspects of texts and readers; this is surveyed below.

An Inappropriate Statistical Model

Arguments against readability formulas are sometimes treated as though they had already been crushed by the weight of accumulated evidence. It is true that formulas can account for as much as 60 to 80% or more of the variance in student responses measures of the ease or difficulty of texts, but the weightiness of this evidence is an illusion. The problem with formulas is that, without any exception of which we are aware, readability researchers have analyzed their data using the wrong statistical model, one in which data are aggregated by grade. This is a problem because almost all users of formulas--for instance, teachers and librarians--are attempting to match books to individuals, small groups within a class, or, maybe, the collection of individual students at a certain grade level in a specific school. For example, a group consisting of students reading between the second grade level and the sixth grade level might have an average level of fourth grade, but a fourth grade level text (also averaged over sample passages) would not necessarily be suitable for each individual student.

In studies such as Vogel and Washburne (1928) and Bormuth (1966) in which readability formulas were validated, texts of a very wide range of difficulty were investigated. Of course, the wider the range of text difficulty the higher the correlations of text features with the student response measure. However, such correlations are unrealistic since a seventh grade teacher, for

instance, will not be considering high school physics texts or first grade primers. When Rodriguez and Hansen (1975) replicated Bormuth's (1966) study using seventh grade students and texts appropriate for seventh graders, they found that the text features accounted for only 20 to 40% of the variance in the student response measure, instead of the 80 to 85% in the original Bormuth study.

It is well-known that aggregating data leads to a big increase in the percentage of variance that is apparently explained. But when formula authors aggregate while users individuate, the increase in variance explained is misleading. The user is left with an inflated impression of the power of the formula to predict the difficulty of texts for individual readers.

The correct approach would be to analyze the total variance, treating both texts and individuals as random variables. This research remains to be done. If it were done, we would not be surprised to find that the best formulas explained, say, 10% of the variance [of individual scores] instead of 80% of the variance [of grade-level averages].

Reading is now understood to be an interactive process (see chapters in Spiro, Bruce, & Brewer, 1980). What this means for readability research is that there should be interactions between characteristics of texts and characteristics of readers. Detecting interactions of this type is impossible when data are aggregated. Moreover, if such interactions do exist, this would mean that a formula that gave a seemingly good prediction of

grade-level averages could be grossly inaccurate when used to select material for any individual reader. The sections that follow summarize evidence showing several strong interactions between text characteristics and reader characteristics and suggest other probable interactions that have not yet been documented in empirical studies.

To encapsulate our conclusion, because an inappropriate statistical model has been used, the right unit for assaying the weight of the evidence from readability research is the ounce instead of the ton. Unless a formula were to include terms representing interactions, not only among text features, but also between text features and reader characteristics, it could not do justice to comprehension as we now understand it.

Correlation is not Causation

In this section, we survey research which has sought to determine what effect word and sentence difficulty has on comprehension of texts. We conclude that these factors, which enter into all formulas, do not directly influence comprehension very much. If their inclusion in formulas is taken seriously as a model of text comprehension, incorrect predictions will be made.

Word Difficulty

The major variable in every readability formula is some operational definition of word difficulty, such as the percentage of words that do not appear on a list of words familiar to children, the length of words in syllables, or the length of the words in letters. It may seem intuitively obvious that long,

rare words are an important cause of text difficulty, but close analysis shows that this intuition is open to serious question.

Nagy and Anderson (1984) have estimated that there are about 240,000 words in printed school English. About 139,000 of these are semantically transparent derivatives or compounds, that is, words that a person could figure out from knowledge of the parts with little or no help from context. Below are several examples, along with the frequency with which each word occurred in the 5,088,721 word corpus that formed the basis for the American Heritage Word Frequency Book (Carroll, Davies, & Richman, 1971):

unladylike	2
girlish	0
rustproof	2
distasteful	4
helplessness	4
caveman	1

For comparison's sake, consider that people occurred 7,989 times in the corpus or that sentence occurred 3,122 times.

Though not all derivatives and compounds are as easy as the ones above, these examples do illustrate the fact that long, rare words are not necessarily, or even usually, hard words. An estimated additional 43,000 words in printed school English are semantically opaque derivatives and compounds. In most of these cases, the word parts provide guides to pronunciation and partial clues to meaning. Some examples are: apartment, saucepan, shiftless, and foxtrot.

Nagy and Anderson (1984; see Table 6, p. 320) found that semantically transparent derivatives are disproportionately found in the lower end of the frequency distribution, far more often than morphologically basic words (words that cannot be divided into parts with consistent meanings) and semantically opaque derivatives. Only 10% of the most frequent words in printed school English are transparent derivatives. As one moves downward in frequency, however, the proportion of transparent derivatives increases steadily, until among the least frequent words there are nearly twice as many transparent derivatives as there are basic words and opaque derivatives.

Thus, most long, rare words are derivatives and compounds, and the great majority of these are phonologically and semantically transparent. What inference can be drawn from this fact about the extent to which long, rare words are a cause of text difficulty? We present evidence below suggesting that they are not a cause of difficulty for most readers. Our conjecture is that these words are a cause of difficulty only for a special subclass of readers, those who are poor decoders, specifically those who have trouble segmenting words into useful parts such as basic words, prefixes, suffixes, and syllables (and perhaps into parts whose status is more problematical such as bound morphemes and phonograms, in the case of words like raspberry, caterpillar, and minister, which cannot be analyzed into meaningful units, even though they might appear to be made up of separate parts).

Most children are able to deal with words productively composed of parts. One of the best established and most

interesting findings of developmental psycholinguistics is that preschool children overextend the rules of inflectional morphology (Berko, 1958; Cazden, 1968). At one time or another, most children three or four years of age can be heard to say, for instance, foots instead of feet or eated instead of ate. Far from indicating that they don't yet know English, these overextensions are a sign that the children are making crucial inductive generalizations about word composition.

Recently, we have uncovered preliminary evidence that knowledge of derivational morphology develops later than knowledge of inflectional morphology. Anderson and Freebody (1983) gave fifth graders a checklist vocabulary task in which real words varying widely in familiarity were to be discriminated from close-to-English nonwords. The fascinating finding was that almost all of the false alarms of the good readers were with "pseudo-derivatives," where a pseudo-derivative was defined as a letter string that does not occur in English, but which consists of a real word and suffix. Among the top quartile of readers, for instance, who checked an average of only 6.4% of the nonwords, 70% checked loyalment, 48% checked conversal, and 19% checked forgivity. Anderson and Freebody (1983, p. 254) characterized these good readers as "aggressive" in applying morphological principles to attack the meanings of unfamiliar words. Notice that, whereas the checklist task in a sense tricked the children into making mistakes, aggressiveness in using morphology would be highly functional during normal reading.

Findings from research in progress suggest that overextensions of the type just illustrated (involving neutral suffixes like -ness that attach to stems with no shift in pronunciation or spelling) peak at about the sixth grade (see Tyler & Nagy, 1986). Fewer overextensions encompassing pseudo-derivatives are observed with fourth graders, presumably because generalizations about derivational morphology are fragmentary among most children at this level. Further, overextensions are no more frequent among eighth graders than fourth graders; presumably at this level, though, eighth graders have learned more of the sometimes subtle selection restrictions on the use of derivational suffixes. Just as the young child eventually learns that you say ate instead of eated so, too, it is reasonable to suppose, does the typical eighth grader tacitly know that forgivity is not right because -ity attaches only to adjective stems of latinate origin.

The tentative conclusion we draw from the foregoing is that for the child in the fifth or sixth grade making average, or even somewhat below average progress in reading, the lion's share of long, infrequent words do not cause increased text difficulty. We do not believe that the typical child able to read at this level would have any more than the slightest problem with even previously unencountered transparent compounds and derivatives, provided the base word or words were known. Of course, long, infrequent words may cause problems for, perhaps, the bottom quartile of middle grade readers, because they cannot reliably decode the words and segment them into useful parts, and probably

have a shaky command of derivational morphology. For similar reasons, long, infrequent words can be expected to cause problems for a larger proportion of children in the primary grades.

We turn now to words that are really difficult for children, not unladylike and helplessness, but rambunctious, tort, or buffoon. Do words such as these cause texts to be difficult? Available research bearing on the answer has yielded weak and inconsistent results. First, there is the readability research, discussed below in this paper, showing that splitting long sentences and substituting short, frequent words for longer, less frequent words generally produces little improvement in text comprehension.

Better evidence, in principle at least, comes from studies in which children were taught truly difficult words and then tested to see whether comprehension of texts containing the difficult words improved. Several studies of this kind have produced non-contrastive 'flat' results. For instance, Jenkins, Pany, and Schreck (1978) explored several methods for teaching the meanings of 12 difficult words. All the methods were at least somewhat better than no instruction. The most effective method with both normal and learning-disabled children involved intensive drill and practice on the words in isolation. However, even when children had definitely learned the meanings of all the difficult words, they did no better than uninstructed children, who definitely did not know the words on a cloze test or in retelling a brief story that contained the difficult words.

That instruction in difficult vocabulary can produce improvement in text comprehension has been demonstrated by Beck and her associates (Beck, McCaslin, & McKeown, 1980; Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson, & Perfetti, 1983). They hypothesized that instruction on difficult words will improve comprehension only if the words are learned thoroughly, so that the word's meaning can be accessed automatically, and so that the word is embedded in a rich mental network of associations. In two studies, involving 75 half hour lessons over a five-month period, during which fourth graders encountered 108 difficult words--such as glutton, filch, lurch, and jovial--10 to 40 times in a range of cleverly designed instructional activities, Beck and her colleagues did find significant increases in comprehension of texts loaded with the words that had been taught. Thus, the hypothesis was confirmed, though the fact that it took such an heroic effort ought to give pause to advocates of direct vocabulary instruction.

A different tack for assessing the influence of difficult vocabulary is described in Freebody and Anderson (1983a). They compared the comprehensibility of nine sixth grade social studies texts containing fairly easy vocabulary with alternate versions of the same texts in which either one-sixth or one-third of the content words were replaced with more difficult synonyms--for instance, descending for falling, pulverize for grind, flora for plants, and minute for tiny. In this study, and three other studies (1983a, Experiment 2; 1983b) in which one-quarter of the words in several texts were replaced, vocabulary difficulty

accounted for an average of only 4% of the variance in three measures of text comprehension. Freebody and Anderson (1983a, p. 36) concluded "that it takes a surprisingly high proportion of difficult vocabulary items to create reliable decrements in performance."

The properties of words and texts that influence the incidental learning of word meanings during normal reading were investigated by Nagy, Anderson, and Herman (1987). Twelve passages, including both expository and narrative texts, were selected from textbooks at the third, fifth and seventh grade levels. The passages contained 212 difficult "target" words (words which would be tested later) judged to be unfamiliar to most children, which were read by a total of 352 third, fifth, or seventh graders. Word properties examined included length, morphological complexity, part of speech, conceptual difficulty, and the strength of contextual support for each word. Text properties included readability as measured by four standard formulas and several measures of the density of difficult words.

Among the word properties, only conceptual difficulty was related to learning the target words. A word was defined as conceptually difficult if the concept associated with it was judged as not known by children in a certain grade, and learning the concept required new factual information or learning a system of related concepts. For example, the noun divide, in the sense of a boundary between drainage basins, cannot be learned apart from other concepts about river systems.

Among the text properties, learning from context was most strongly influenced by the proportion of target words that were conceptually difficult and by the average length of target words. These two variables, both of which suppressed learning, were fairly highly intercorrelated, but appeared to contribute independently to predicting word learning.

Interestingly, none of the readability formulas applied by Nagy, Anderson, and Herman significantly predicted the learning of word meanings during reading, unless the proportion of conceptually difficult words entered the equation in a multiple regression analysis. This variable accounted for 4% of the variance. Before it entered, the four readability formulas accounted for an average of 1% of the variance; after it entered, they accounted for an average of 2%.

In summary, word difficulty does not seem to be as important a direct cause of text difficulty as might be assumed looking at readability formulas. First, most long, infrequent words are transparent derivatives and compounds that would not be expected to be difficult for the typical student by the time he or she reaches the middle grades. Second, whether or not a transparent derivative or compound is actually difficult for a particular child will depend upon the child's level of understanding of derivational morphology and on even more basic abilities in decoding and segmenting words. Hence, this is clearly one of the cases where interactions are expected, and where it can be anticipated that formulas fit to grade-level averages will do a poor job of predicting individual understanding. Third, even

words that readers definitely do not know do not appear to cause big problems in comprehension, unless the text is dense with such words, and the words meet strict criteria of conceptual difficulty. Fourth, as an inference from the foregoing, the prominent role that measures of word "difficulty" play in readability formulas probably means that the measures are largely indirect reflections of the deeper factors that cause comprehension difficulty. To preview the argument that will be developed in a later section, a text with a lot of unfamiliar words is usually about an unfamiliar topic, and it is mainly lack of knowledge of this unfamiliar topic that makes comprehension difficult.

Finally, we cannot resist the observation that after 60 years of research and an estimated 1,000 or more books and articles (Klare, 1984), an adequate and theoretically defensible analysis of word difficulty, the principal variable in every formula, has not heretofore issued from readability research. We attribute this embarrassing fact to shallow empiricism arising from a preoccupation with what "works."

Sentence length. No recent study has focussed specifically on the contribution of sentence length per se to comprehension. Preliminary findings from an as yet unpublished study by Davison, Wilson and Hermon show that sentence length alone accounts for a very small percentage of the variance in the comprehension of texts. Average sentence length is correlated with complexity of internal clause structure, which in turn is correlated with the presence of markers of subordination and of connectives (so, or,

because, when if, and even and, etc.) which make explicit the meaning relation between clauses. Hence, long sentences usually consist of syntactically connected clauses with conjunctions or other markers of connection. The results of the study of seventh grade readers by Davison, Wilson and Hermon suggest that texts with long sentences are comprehended as well as short sentences, except by poor readers, those in the bottom third of students at this grade level.

Connectives in sentences are not necessarily what makes a long sentence difficult. There is a body of evidence which suggests that, far from being a source of difficulty, the presence of conjunctions facilitates comprehension, particularly when two clauses could be connected in more than one way, such as in a 'reversible' way. For example, the two sentences in (1) may bear more than one relation to one another. These different interpretations are paraphrased in (2a) and (2b), in which an explicit connective is used.

1) I moved the switch. The lights went off.

2a) I moved the switch, because the lights went off (to turn them back on).

2b) The lights went off because I moved the switch (turning them off).

If there is no connective, the reader is not always able to make the correct inference, especially if it is not clear from the context which inferences (if any) should be made. In another example, the two sentences in (3) can convey two very different meanings, (4a) and (4b).

3) Let's fill the bird-feeder with seed. The cat hasn't been active lately.

4a) Let's make the cat more active by filling the bird feeder.

4b) It's safe to fill the feeder because the cat isn't active.

The presence of explicit connectives is often helpful to the reader if the context does not make sentence connections obvious.

Pearson (1974-75) has shown that children prefer sequences of sentences containing an explicit connective such as because, and understand them better than sequences of short, implicitly connected sentences. Irwin (1980) showed that for somewhat longer texts both fifth graders and college students comprehended reversible causal relationships among sentences better if an explicit conjunction was used. In a subsequent study, Irwin and Pulver (1984) found that for fifth and eighth grade students, comprehension of reversible causal relationships was improved if the conjunction was explicit, and not simply left to be inferred. The presence of a conjunction thus facilitates comprehension, even though it adds to average sentence length in the text. A conjunction affected students independently of reading ability. If sentence length is a factor in comprehension, it would be expected that longer sentences would pose a greater problem for students who are poor readers than those with better reading ability. Irwin and Pulver found no interaction between sentence length and reading ability, however.

Increases in sentence length do not necessarily impede understanding. Beck, McKeown, Omanson & Pople (1984) systematically revised two basal reader stories to improve comprehensibility. The revisions were directed at eliminating difficult surface forms, such as pronouns with unclear antecedents; clarifying references to concepts the readers might not know; and clarifying relationships among parts of the story. In recall of the central elements of the story, both skilled and less skilled third grade students did better after reading the revised versions, even though the readability level was raised one grade level on the Fry scale by the revisions.

A study of adults' comprehension of difficult and unfamiliar material by Charrow and Charrow (1979) compared a revision of the jury instructions written following the implicit guidelines of readability formulas, to one written according to a set of guidelines based on psycholinguistic research and a careful analysis of the content of the instructions. One set of revisions was done by simplifying words and shortening sentences, so as to decrease the readability score computed for the passages. These revisions, which aimed at lower readability scores, resulted in no greater recall than the original forms, and in some cases even poorer recall.

The other set of revisions focussed on the important pieces of information in the instruction, eliminating distracting less important phrases and drawing attention to the central concepts. The language was revised to make the sentence structures match the content more clearly, and to use passive, embedded and

preposed structures only when they were supported by the surrounding context. For example, compare the original and revised versions of part of the definition of contributory negligence:

5a) (original)

An essential factor in contributory negligence is that it contribute as a proximate cause of the injury. (Charrow and Charrow, p. 1354) (17 words)

5b) (revised version)

If the plaintiff was contributorily negligent, he actually helped cause his own injury, through his own negligence. (Charrow and Charrow, p. 1355) (17 words).

Here, clarifying sentence structure and vocabulary caused increased comprehension. Nevertheless the sentences in (5a) and (5b) are the same length, and the vocabulary in both cases is technical and infrequent. The revisions of the type illustrated in (5b) were not much different in readability level from the originals, but they significantly improved the subjects' ability to recall and paraphrase the instructions.

In this next section we will discuss some cases in which comprehension of a sentence is made more difficult by some features of the sentence itself. We will show, however, that difficulty of comprehension is not linked in a simple way to complex features of sentence syntax. That is, complex features of sentence structure do not necessarily present a problem every time they occur. For example, if the context fits the complex structure and justifies its use, the structure may not be

difficult to comprehend. But in other cases, there may be a mismatch between the features of a sentence and the context in which it occurs, and in that case, it may well be difficult for a reader. Or if processing a complex structure in some way exceeds the attentional resources of the reader, it will be difficult. As we will see, difficulty of sentence structure is not an absolute value, and depends on interactions with other text features and with features of the reader.

The sentence length variable may reflect some kind of semantic complexity in the text, but as we have seen in the studies just reviewed, there is no general causal relation between how long a sentence is and how easy it is to understand. This is not to say that sentence structure has no effect on how well a sentence can be understood. It is easy to imagine many ways in which the length and complexity of a sentence could make it hard to understand, and conversely, how sentences may be written so as to make their meaning easy to understand. What is not easy to characterize is some general definition of sentence complexity, because this is not an absolute value. Specific sentence features do not always introduce difficulty into the processing of the sentence that contains them. Sentence features interact with other sentence features, and with features of readers, in many cases where difficulty of comprehension has been revealed by experimental measures, as in the Irwin and Pulver study (1984) cited earlier.

A long sentence may be hard to understand simply by virtue of its length, all other things being equal, just because it

contains a large number of words to identify and access. But if we compare sentences of exactly the same length, with the same words, we may find that they differ in complexity. For example, Irwin and Pulver used sentence pairs like the following:

6) Because Mexico allowed slavery, many Americans and their slaves moved to Mexico during that time.

7) Many Americans and their slaves moved to Mexico during that time, because Mexico allowed slavery.

The subjects, who were asked to answer comprehension questions about these sentences, were third, fifth, and eighth grade students, as well as college students. As noted earlier, versions of the sentences with connectives, though longer, were understood better than the single clause sequences. What surprised the experimenters, however, was that the version with the preposed adverbial clause, (6), was difficult for the younger subjects, those in the third and fifth grades. They predicted that (6) would always be <u>easier</u> than (7) because the order of the clauses puts cause before effect, and this is generally preferred. Older and more skilled readers had no trouble in matching the order of mention with the meaning of <u>because</u>. But, apparently the younger and less skilled readers did not use the cause-effect ordering in the same way and could not overcome the difficulty they had in understanding the sentence structure.

Why should a preposed clause be more complex than a similar clause which follows the main verb and its objects? A very broad explanation comes from work by Yngve (1960), who wanted to define what is involved in producing or understanding a sentence. The

parts of a sentence consist of words grouped into smaller and larger phrases, belonging to different categories whose features are defined by the rules of the language. For example, words like the occur only in phrases with nouns and precede the noun. This word is a left branch within a Noun Phrase, and its appearance signals the beginning of a phrase of the NP category. Hence it is stored in working memory while the next constituents are searched for, including the noun. Yngve proposed that for this reason, left branches always require more memory capacity to produce or understand than right branches. Preposed adverbial clauses are left branches, large phrases which must be held in working memory until the main clause constituents are found (Bever & Townshend, 1979).

Kemper (Kynette and Kemper, to appear) investigated people at the other end of the age range than in the Irwin and Pulver study, elderly adults who have begun to have less working memory capacity than younger adults. She compared their ability to paraphrase or recall sentences with left branching or right branching structures. The sentences in (8a) - (10a) all have left branching structures, while those in (8b) - (10b) have right branching structures.

Free relative clauses:

8a) [What I did] interested my grandchildren.

8b) My grandchildren watched [what I did].

Finite that clauses:

9a) [That the cookies were brown] surprised me.

9b) I believed [that the cookies were brown].

Relative clauses modifying noun phrases:

10a) The cookies [that I baked] were delicious.

10b) My children enjoyed the cookies [that I baked].

In a study of journals written over a span of many years, Kemper found that the writers produced very few left-branching structures of these types as they became elderly, compared with middle age. She also found that elderly adult subjects had more trouble paraphrasing sentences with the left-branching structures than the right-branching ones. In another study, the subjects, when asked to read connected texts, recalled fewer left-branching structures than their right-branching counterparts. Interestingly, the subjects had less difficulty with left-branching sentences when they expressed the most important information in the passage. This is another instance of an interaction within a passage.

Under some conditions, then, left-branching structures appear to be more complex than right-branching structures. Nevertheless, there have been numerous objections to Yngve's general proposal that left branches always introduce complexity in the position in the sentence where they occur (for a general discussion see Frazier (1984)). For example, sentences like (11) are read no differently than sentences like (12), according to the eye-movement data in Frazier, Rayner, and Carlson (ms, cited in Frazier, 1984):

11) [That the traffic in this town is unregulated] bothers me.

12) It bothers me [that the traffic in this town is unregulated.]

If a pronoun occurs in the embedded clause, however, sentences of the type in (13) were read more slowly than those in (14):

13) [That people look at him strangely] bothers Mary.

14) It bothers Mary [that people look at him strangely].

The young adult subjects in Frazier's study had difficulty with a left branch only if there was an additional relation such as anaphora to be processed at the same time.

A single left branch structure is not as difficult to process as multiply embedded ones, as in (15):

15) That that men were appointed didn't bother the liberals wasn't remarked upon by the press. (Frazier (1984:163)).

Frazier (1984) speculates that the correct interpretation of such a complex sentence requires a great deal of abstract (and left-branching) structure in proportion to the number of words in surface structure. This amount of structure containing internal sentence phrase nodes overloads temporary processing capacity. Frazier reports that sentences like (16) appear to many readers to be well-formed, even though one verb phrase is missing:

16) That that men were appointed didn't bother the liberals. (Frazier (1984:179)).

The first <u>that</u> needs to be matched with a predicate (e.g., <u>wasn't reported</u>), whose subject is the internal sentence <u>that men were</u>

appointed didn't bother the liberals. To detect this anomaly requires that a lot of structure be kept in working memory, too much even for most normal adults.

Even complex structures like these are not absolutely difficult to process. The presence of conjunctions with specific syntactic properties and semantic content makes it easier to understand sentences like (12) and to detect missing phrases (cf. Frazier, 1984:178-80).

17) Since if you light a match the gas will explode, you should be careful.

This sentence contains two left-branching structures, one nested within the other. It is nevertheless not as difficult to understand as (15), which has the same general structure.

Though some sentences like (15) are harder to understand than others like (17), it is not always clear what makes the difference. The hypothesis, however, is that left-branching structures may cause an overload on working memory, with resulting problems of comprehension, if the reader has some problems with short-term memory, as very young or very old readers may. People with normal capacity may also have problems with left-branching structures if some other factor makes demands on short-term memory and there are no additional surface cues which add information. The tendency of left-branching structure to make a sentence hard to understand results from an interaction between the demands on short term memory caused by left-branching structures and a number of other factors.

Yngve's proposal that left-branching and deeply embedded structures are complex has been used to construct a predictor of complexity, which automatically assigns weightings to syntactic structures from which a complexity profile could be derived for a whole sentence or text (Botel & Granowsky, 1972, and Botel, Dawkins, & Granowsky, 1973). While this approach is interesting, it was never pursued in detail at the time it was proposed nor used to make specific predictions tested with comprehension measures. Perhaps if it had been, there would have been some alternative conceptions to readability formulas. If sentence complexity is the product of interactions rather than an absolute value, however, it is still unlikely that refinements of the formulas to measure sentence complexity would have led to more accurate predictions.

Another attempt to refine the measure of sentence complexity was in the form of a taxonomy of structures which seemed to be acquired late in childhood or to cause difficulties in comprehension for young children, according to psycholinguistic studies of language acquisition and comprehension in the 1960s and early 1970s (Dawkins, 1975). There are several problems with this approach. First, more refined experimental methods have shown that children can understand complex structures at an earlier age than previously thought. For example, Sheldon (1974) reported that young children interpreted restrictive relative clauses like (18) as though they were conjoined structures describing successive events (19):

18) The dog which bit the cat ran away.

19) The dog bit the cat and ran away.

But Hamburger and Crain (1981), found that if sentences are placed in a natural discourse context, young children correctly understand a sentence like (18) as a way of picking out which of several dogs is being referred to.

Second, the complexity of a particular construction like the passive or relative clauses does not always cause it to be difficult to understand. It is hard to imagine why a language has both an active and a passive form for clauses unless there is some difference in their functions. It would be strange if the only use for passive clauses was to express information in a more complex or obscure way than in active clauses. In fact, as many experimenters have shown (Glucksberg, Trabasso, & Wald, 1973; and Olson & Filby, 1972; for example), passive sentences require less reading time and are more accurately comprehended when the preceding verbal context contains an antecedent for the passive subject, which is the topic of the target (passive) sentence.

The relation between syntactic features of a sentence and the topic is discussed in relation to context in Davison and Lutz (1984) and Davison (1984). The two sentences in (20) differ in that the subordinate clause subject in (20a) has normal subject properties, while the corresponding word him in (20b) is semantically a subject, but has properties of an object.

20a) We believe that he is intelligent.

20b) We believe him to be intelligent.

The constituent him in (20b) is like the subject of a passive sentence, since him has the syntactic markers of one grammatical role and the semantic properties of another role. So if we assume that sentence structures are more complex if the outward markers of grammatical roles do not directly correspond to the semantic relations, the structure in (20b) is more complex than the synonymous structure in (20a).

The difference can be seen by placing the more and less complex versions of a sentence in a discourse context. For example, consider the sentence (21) to be the context preceding either (22a) or (22b):

21) People are afraid to go out at night.

22a) We believe that a flying saucer is exploring Chicago.

22b) We believe a flying saucer to be exploring Chicago.

The subordinate clause subject a flying saucer in the second version (22b) is more like an object. The sentence fits this context less well than the less complex version (22a). There is some lack of continuity between (21) and (22b), as though the existence of a specific flying saucer has to be assumed, although it had not been mentioned. For (22a), there is no such assumption conveyed. In the case of (21) - (22b), however, the reader must make an inference linking the two sentences, in somewhat the same way as when the definite article the is used (Haviland & Clark, 1974). The difference in discourse continuity originates in the difference of sentence structure. It appears, then, that there is an interaction between sentence structures and the context in which the sentence occurs. If the context

contains discourse antecedents for certain phrases which the syntax marks as special, then the more complex structures are not necessarily harder to understand. In fact, the more complex structures may facilitate comprehension by showing how the new sentence is to be linked to the context. Complexity may arise only when a linguistic form like do so requires a matching structure in a previous sentence, and none is found (Tanenhaus & Carlson, 1985).

There is also an interaction between complex words and difficult syntactic structures. Complex words like indecisive and indecision have a transparent structure, so that their meanings are composed from their parts. Part of their structure includes a suffix which marks the syntactic category of the word, -ive for an adjective and -ion for a noun. Tyler and Nagy (1985) found that some subjects may ignore this information in the understanding of certain types of sentences, even when they correctly use the words in another task. In sentences like (23) and (24), the suffixes in indecisive and indecision are associated with quite different sentence structures:

23) People were afraid of a general indecision about nuclear war.

24) People were afraid of a general indecisive about nuclear war.

The subjects in Tyler and Nagy's study chose the paraphrase appropriate for (23) as the preferred interpretation for both (23) and (24), ignoring the adjective suffix -ive which makes this interpretation inappropriate for (24). The reason seems to

be that the sentences are ambiguous between two syntactic phrase structures up to the point where the target word appears. Parsing strategies which tend to maximize the choice of the simpler interpretation lead to a preference for the interpretation [$_{NP}$ a general N . . .] rather than the more complex interpretation [$_{NP}$ [$_{NP}$ a general] [Adj...]]] (cf. Frazier & Fodor (1978)). These parsing strategies lead to a syntactic decision about the phrase structure of the sentence before the target word is encountered. If we assume that abandoning a decision which is already made and reprocessing the sentence adds to complexity of processing, then it is not surprising that the initial choice for N is retained, even when the word has adjective features. So even someone who can normally make use of the information in affixes may ignore it in the face of other factors which add to the complexity of the sentence being understood.

In this section we have discussed a number of cases in which syntactic features of a sentence may make the sentence difficult to understand. But the complexity which is introduced is the result of the interaction of several factors all being processed at once in some limited space in working memory (as we will note in the section which follows). The features of sentence structure cannot be used as absolute indicators that the sentence will be complex, so that it is not possible to replace the length measure with some other direct measure of complexity, however detailed and sensitive it might be. What is measured in this way might pose a problem for some readers if other factors are

present. While there are explanations for why some sentence features may overload processing capacity in some cases, we are a long way from a general characterization of sentence complexity and how it arises.

Sentence length and word complexity are measured in a sample of text in computing its readability. These variables do not, however, directly reflect the properties of a text which make it difficult for a reader to read and comprehend. As is well-known, the application of a formula in reverse, revising a text to make the sentences shorter and the words simpler, does not increase comprehension. The complexity of a text may be directly indicated by the linguistic factors which are measured by formulas. The studies just cited show that the same factors, complex morphology and sentence connectives, actually convey information about meaning in an explicit way and so are not barriers to comprehension for most readers. They may appear to be powerful indicators of complexity because of the inappropriate use of an aggregate statistical model, which does not take into account the interaction of properties of the individual with other properties of the text. In the next section we discuss how some of these other factors, not measured by formulas, have a direct influence on comprehension.

Limitations on Processing Capacity

Thus far, we have presented evidence and arguments that point to the inescapable conclusion that readability formulas permit an exaggerated impression of the role of word difficulty and sentence complexity in text comprehension. However, it would

be foolish to suppose that these elements of language have no influence on comprehensibility.

Connected written text has many features, including content, style and organization. But at the most basic level it is composed of words organized into sentences, which conform to the grammatical rules of the language in question. Ultimately it must be interpreted on that level, so that the text as a whole must pass word by word and sentence by sentence through the 'bottleneck' of the linguistic processor, in the metaphor used by Perfetti and Lesgold (1977). The comprehension of words and sentences requires linguistic knowledge which is not wholly or even largely predictable from contextual factors. The meaning of complex expressions is composed from the meaning of the parts and the ways they are put together, according to the rules of the language. The ability to understand a text at this fundamental level requires linguistic knowledge.

Words and sentences in a text are the raw material entering into a 'full' interpretation which is only partially determined by the words and sentence meanings. These meanings enter into higher level cognitive processes such as making inferences, combining propositions about the same referent, and integrating propositions with knowledge which the reader already possesses. If, as we have shown, linguistic factors do exert some influence on how difficult a text may be for a reader, we need to relate word difficulty and sentence complexity to a sound model of how language is processed.

If some features of words or sentence structure delay comprehension, or simply make it more difficult, the influence of these factors will not necessarily be reflected in failure to answer comprehension questions correctly. The ability to answer such questions will be based on an interpreted representation of meaning, perhaps combining the meaning of a specific sentence with other information. Even cloze questions, which consist of gaps in texts, are answered after the surrounding sentences have been interpreted. Answering comprehension or cloze questions, therefore, is based more on a memory of representation of a sentence than on a sentence piece by piece while it is being processed.

The linguistic form of a sentence is not always available after it has been stored in memory. In two studies which have strongly influenced conceptions of language interpretation, Bransford, Barclay, and Franks (1972), and Bransford and Franks (1971) showed that subjects do not always recognize a sentence in excactly the same form in which it was presented; instead, they reliably remember the meaning of a sentence but not its exact surface form. It appears that once a sentence has been interpreted, it is usually no longer necessary to retain a representation of its form. To do so would require extra memory resources. It appears from Jarvella's classic study (1971) that working memory resources are used very economically. If subjects are interrupted while reading and asked to decide if they have seen a certain word before, they can make this decision much more rapidly if the word occurred in the clause currently being read

than if it occurred in a previous clause or preceding sentence. Assuming that retrieval from current working memory is faster than from longer-term memory, it appears that sentences are processed in chunks the size of a clause or possibly smaller (Marslen-Wilson, Tyler, & Seidenberg, 1980).

Marslen-Wilson's (1975) finding that syntactic or semantic errors are very rapidly detected and corrected also shows that processing of oral language is extremely rapid, and the same must be true of written language, at least for fluent readers. While many important details are unclear, a model of language processing which is consistent with these findings assumes a temporary working memory with a limited capacity which has the function of breaking a linguistic input into chunks and applying lexical and other linguistic knowledge to the chunks to derive an interpretation. This interpretation, whose form is not directly observable, lacks some, if not all, features of surface structure. As a meaning representation of the sentence is constructed, it is stored in long-term memory and can be combined with other semantic material.

The best time to look for the influence of linguistic factors on language understanding is at the moment of processing, rather than after the interpreted meaning of the sentence has been stored, and, hence, already subjected to reinterpretation or revision from other information from the text or background knowledge. For this reason, the measures used in experiments where linguistic factors are a variable tend to be either those very sensitive to details of comprehension, such as immediate

recall, or on-line measures which are sensitive to direct loads on attention and processing capacity. These measures include reading time for specific words or sentences, decision time and accuracy for tasks which immediately follow reading or recordings of the fixations and movements of the eye (cf. Frazier & Rayner, 1982).

To the extent that readability formulas measure factors of sentence and word complexity which have some direct influence on comprehension, they are crude approximations of a model of processing capacity. Studies reviewed in earlier sections showed that some complex linguistic factors interfere with comprehension, causing difficulty when they place heavy demands on immediate processing capacity. Certain kinds of readers, such as young children or elderly people, are likely to have less immediate processing capacity than others. Other readers have difficulty if they must deal with a great deal of material at one time, though what causes difficulty is not well understood at present since many linguistic factors may interact either to cause or to mitigate and remove processing difficulty. Perfetti and Lesgold (1977), among others, argue that word decoding places a very heavy burden on processing capacity in poor readers, such a heavy burden that either resources are exhausted for higher level processing, or the scheduling of the processing operations is disrupted. This is a promising hypothesis which needs to be understood in more detail, as do other cases where interactions of different factors influence comprehension.

This is also the case for factors which improve comprehension, such as interest and rich background knowledge (see below). Do these features of the reader in conjunction with the text somehow increase processing capacity for the initial interpretation of the linguistic material? Or do they increase the efficiency of higher-level processes, leading to fewer wrong inferences, more direct interpretation of anaphoric relations, better integration with material in the context? Or does interest simply increase the reader's motivation to go through the processes of interpretation, making best use of whatever capacity to understand language which he or she may possess? Not very much is known about these issues or about how good and poor readers differ, if they do, in general knowledge of language, as opposed to decoding and other processes specific to written language (cf. Perfetti & Lesgold, 1977).

While much remains to be investigated, it appears to us that the issues discussed above are far more promising questions to pursue than those asked in traditional studies associated with readability and readability formulas, which are concerned with statistical correlations, ease of application and "what works." These studies have sought to show greater or lesser correlations of comprehension measures with linguistic variables as measured in various ways. The strongest predictors of comprehension, measured retrospectively with comprehension or cloze questions, have always turned out to be sentence length and word complexity, which are not truly independent of one another, in any case. While these studies may satisfy short-term goals, they do not

reveal anything of interest about the functioning of cognitive processes applied to understanding language. They do not illuminate why a text is difficult to understand, or how comprehension is affected by interactions of features in the text, the language and the reader. We turn now to some other aspects of texts which affect comprehension.

Prior Knowledge

The knowledge a reader already possesses about a topic exerts a powerful influence on comprehension of texts about that topic. This has been demonstrated with readers of every age and all manner of topics. A sampling: Pearson, Hansen, and Gordon (1979) found that second graders who knew a lot about spiders comprehended more from a text about spiders than second graders who were comparable in IQ and reading level but knew little about spiders. Spilich, Vesonder, Chiesi, and Voss (1979) asked college students high and low in knowledge of baseball, but equivalent in verbal ability, to read and recall a story about a half inning from a fictitious baseball game. Those who knew a great deal about baseball, particularly information of tactical significance to the game, recalled more information than those who knew little. Sticht, Armijo, Weitzman, Koffman, Roberson, Chang, and Moracco (1986) showed that Navy personnel with high scores on a test of Navy technical knowledge could comprehend Navy texts five grade levels higher, as determined by the Flesch-Kincaid formula, the formula officially prescribed by the Navy, than personnel with low scores on the test of knowledge.

Comprehension will vary, depending upon the match between readers' actual knowledge and the knowledge presupposed by texts. This has also been demonstrated a number of times. For instance: Steffensen, Joag-dev, and Anderson (1979) had natives of India and the United States read and recall letters about an Indian wedding and an American wedding. Each group read what for them was the native passage text more quickly than they read the foreign text; they recalled more propositions from the native text, especially propositions rated as important by fellow natives; and they introduced more culturally appropriate elaborations of the native text but more culturally inappropriate distortions of the foreign text. In a similar study, Lipson (1983) gave American middle grade Catholic and Jewish students texts about a first communion and a bar mitzvah. Prior religious knowledge strongly influenced their measures of comprehension. Each group read the culturally familiar text in less time, recalled more propositions from it, and made more appropriate inferences and introduced fewer errors when recalling the culturally familiar text. Comparable findings have appeared in research with college students, depending on their major field of study (Anderson, Reynolds, Schallert, & Goetz, 1977), and junior high school students, depending on whether they were black or white (Reynolds, Taylor, Steffensen, Shirey, & Anderson, 1982).

The knowledge a person possesses depends upon age, sex, amount and kind of education, race, religion, occupation (or occupation of parents), hobbies, country of origin and residence,

and region within country, among factors that come readily to mind. Thus, interactions between the knowledge readers possess and the knowledge demands of texts are bound to be the rule rather than the exception, and the complaint made earlier against statistical models in which data are aggregated has more than hypothetical force.

We believe that the reason vocabulary difficulty is the principal component of every readability formula is primarily that it serves as a proxy for background knowledge (see Anderson & Freebody, 1981, and Anderson, Mason & Shirey, 1984, for earlier statements of this hypothesis). This position can be illustrated using words from the Indian wedding text employed by Steffensen, Joag-dev, and Anderson (1979). Only two words in the text, sari and dhoti, would have been unfamiliar to any of the American readers. Neither word figured importantly in the text, so not knowing them could not have had much effect on comprehension. Nonetheless, a test examining knowledge of the two words would have been an excellent predictor of performance. All the Indians would have known both words; some of the Americans would have known sari but few would have known dhoti. It is apparent that the test would have divided subjects in terms of their knowledge of Indian culture, which, of course, was the real reason for the large advantage Indians had on the various measures of comprehension, learning, and remembering.

What we wish to argue is that there is a correlation between the knowledge demands of texts and the use of long, infrequent words and long, complex sentences. We wish to argue, further,

that in made-for-school texts the correlation is higher than any necessity requires. Since the dawn of the readability movement 60 years ago, the heavy controls placed on school texts have made the language in them progressively more simple, unnaturally simple, we believe. In turn, as new readability research has been done, it has fed back in ever stronger form the conclusion that the younger the reader the simpler the language ought to be. The result of generations of inbreeding is, in the words of Anderson, Mason, and Shirey, (1984, p. 35), "that the confounding of knowledge demands and language complexity has been exacerbated . . . [T]he formulas now in use egregiously overestimate the importance of surface features of language. Probably most third-grade students could get the gist of a story about a girl and her puppy even if it were dressed up in fancy language, whereas no amount of simplification of [the language of] an economics treatise would permit very many third-grade students to grasp the concept of the multiplier effect."

Interestingness

As important, or perhaps even more important than the influence of prior knowledge, is the influence of interest on comprehension. In four experiments involving over 400 third and fourth graders, Anderson, Shirey, Wilson, and Fielding (1986) compared the learning and recall of sentences that children find interesting, such as The huge gorilla smashed the school bus with his fist and The hungry children were in the kitchen helping mother make donuts, with ones they find uninteresting, such as The old shoes lay in the back of the closet and The fat waitress

poured coffee into the cup. The newsworthy finding was that interest, as rated by other children, accounted for over thirty times as much variance in sentence recall as readability. It should be emphasized that the sentences were selected so that interestingness and readability were independent and so that there was a wide range of readability. According to the Fry scale, sentence readability ranged from the first to the seventh grade.

Studies using texts have revealed similar, if less dramatic, results. Notably, in a series of well-designed studies, Asher and his associates (Asher, 1979, 1980; Asher & Geraci, 1980; Asher, Hymel & Wigfield, 1978; Asher & Markell, 1974) determined children's interests by having them rate photographs representing a wide array of topics (e.g., ballet, basketball, cats, airplanes, circus). Later, the children read Brittanica Junior Encyclopedia selections on topics that they had individually rated as high or low in interest. Briefly, the findings were, first, that the children indicated far greater desire to read selections on highly rated topics. Second, children's comprehension was superior on high-interest material; in each study, children attained higher cloze scores on their high-interest selections. Third, in two of the studies (Asher & Geraci, 1980; Asher & Markell, 1974), boys' performance was facilitated more than girls' performance by high interest material, a finding since replicated by Anderson, Mason, and Shirey (1984) and Baldwin, Peleg-Bruckner, and McClintock (1985).

A worry is that prior knowledge and interest are not clearly separable. One would suppose that people would be knowledgeable about topics they are interested in, and maybe vice versa. However, Baldwin, Peleg-Bruckner, and McClintock (1985) found only a slight correlation between tests of knowledge of ten topics and interest in the topics among a sample of seventh and eighth graders of above-average ability. They explained this seemingly counterintuitive finding in the following way (p. 502): "[S]chool children . . . are forced to study a variety of topics whether they like them or not. It should not be surprising then to find that a group of above average students could be fairly knowledgeable about space exploration and American Indians, for example, without having any real enthusiasm for those subjects." Baldwin et al. also found that both knowledge and interest independently predicted comprehension of encyclopedia passages on the ten topics.

Systematic empirical study of the features of language, style, plot, characterization, content, and theme that make texts more or less interesting to various readers is in its infancy (for a sampling of work, see Anderson, Shirey, Wilson, & Fielding, 1984; Bettleheim, 1976; Blom, Waite, & Zimet, 1970; Bruce, 1984; Green & Laff, 1980; and Jose & Brewer, 1983). While this field matures, one should not neglect the insights of rhetoricians nor undervalue the craft of skillful writers, as Graves and Slater (1986) have demonstrated in striking fashion. They persuaded three teams of writers to revise a passage from a high school history textbook on the war in Vietnam, described by

one of the teams as "some of the driest prose we had ever had the displeasure of reading."

Graves and Slater's first team was made up of a pair of "text linguists" whose revisions were directed at such matters as clarity, coherence, and emphasis. Below is the material on the Communist guerrillas in the text linguists' revision, which is unchanged from the original except for the addition of the phrase, "in particular:"

> In South Vietnam in particular, Communist forces (the Viet Cong) were aided by forces from Communist North Vietnam in a struggle to overthrow the American-supported government.

The next team consisted of two college composition instructors. In their words, "The six main purposes we had in mind . . . were simplifying information, adding background information, clarifying information, supplying transitions, emphasizing key material, and keeping the passage smooth and readable." Here is what they produced on the guerrillas:

> In South Vietnam, Communist guerillas called the Viet Cong were aided by forces from Communist North Vietnam in a struggle to overthrow the American-supported government.

The last team, a pair of veteran Time/Life editors, revised the passage in a radically different way. In the words of one of them, "To intensify the action, I replaced weak verbs such as 'tried to get,' 'moved,' 'fight,' and 'increased' with words such as 'tried to gain,' 'hustled,' 'grappled with,' and 'skyrocketed.' I added metaphors [and] colloquialisms. . . . However, tinkering with the language did not give the passages a

Time/Life quality: They were still too panoramic, too impersonal. . . . To enrich the content, I inserted 'nuggets' gleaned from library sources. Nuggets are vivid anecdotes and details that remind us that PEOPLE, not events, make history. A Time/Life story is not so much a sequence of events as a string of nuggets. . . . I also quoted from Presidents Eisenhower and Kennedy. After all, why should the textbook quote Kennedy's statement that South Vietnam was of 'vital interest' to the U.S. when Kennedy so graphically called the country 'the cornerstone of the Free World in Southeast Asia, the keystone to the arch, the finger in the dike'?" Below is what this team said about the guerrillas:

> Aided by Communist North Vietnam, the Viet Cong guerrillas were eroding the ground beneath South Vietnam's American-backed government. Village by village, road by road, these jungle-wise rebels were waging a war of ambush and mining: They darted out of tunnels to head off patrols, buried exploding booby traps beneath the mud floors of huts, and hid razor-sharp bamboo sticks in holes.

Groups of eleventh graders read the original passage on the Vietnam War or one of the revisions written by the three teams. They then wrote essays which were evaluated in terms of the percentage of the information in the text that was recalled. The results were that the text linguists' revisions produced a 2% gain in information while the composition instructors' revisions produced a 2% loss. In profound contrast, the Time/Life editors' revisions produced a 40% gain. Informed of their poor showing

and given a second chance to revise the text, the text linguists and composition instructions did better; they produced gains in recall averaging 16% and 21% respectively, while the Time/Life editors held their ground at 37%.

The points that should be made about interest and readability are essentially the same as the points about prior knowledge and readability. First, whether a text is interesting is probably a more potent predictor of its comprehensibility than the surface features of language embodied in readability formulas. Second, readability formulas probably get some of their predictive power because the word difficulty measure is an indirect indicator of whether the text is interesting. Third, there are almost certainly interactions between the topics individual readers are interested in and the stylistic features that please them with the topics and styles of texts; therefore, again, it is dangerous to try to predict individual performance using an aggregate statistical model.

Conclusion

In this paper, we have surveyed the problems arising from treating word and sentence complexity as the direct causes of difficulty in comprehension, and have noted the far greater influence on comprehension of text and reader properties not measured by formulas. We have looked critically at readability formulas from several perspectives. In doing so, we have been concerned with how close these formulas come to being accurate and informative predictors of comprehension, when specific readers read a specific text. In most research on readability to

date, very high correlations are reported between the predictions of formulas based on text features such as word complexity and sentence length and measures of comprehension associated with reading ability. We suggest that these high correlations are the by-product of using an inappropriate statistical model which aggregates texts and readers, and gives an exaggerated impression of the contribution of linguistic factors in the text to ease or difficulty of comprehension. We propose instead that both texts and readers are more appropriately treated as random factors. This approach will lessen the correlations of text properties and predicted grade level, and will also give a more accurate picture of what causes a text to be difficult to understand.

The presence of long sentences and complex words in a text in some way reflects or is correlated with complexities of subject matter, but need not directly cause a text to be difficult. While these factors may impede comprehension for some readers who have difficulty segmenting words and parsing sentences or who have limited working memory capacity, these very same factors also provide the reader with explicit information about the composition of a word or the relations between sentences.

Recent research in reading and the perception of language has used more sensitive measures of comprehension than those which were previously used, either for overall comprehension of whole texts or for the processing of specific parts of a sentence in working memory. These new measures have made it possible to see in more detail what factors interact when a reader interprets

a text. Some of these interactions hold between different linguistic features, and some between the properties of the text and the properties of the reader. Certain kinds of sentences or complex words may be difficult for readers with less processing capacity available in working memory than people usually have. Readers without adequate background knowledge for a text find it much harder to read and understand than readers who have the right background knowledge. A text whose content and way of presenting information are boring to the reader is less well understood than a text which falls within a particular reader's interests.

Clearly, while texts differ in the complexity of the language they are written in, so, too, do readers differ in decoding and parsing skills, background knowledge, and interests. Since reading and understanding a text requires the reader to interact with the text, using his or her knowledge and skills, it is not surprising that there are many factors about readers and texts which cannot be described in terms of a readability formula of the traditional kind. Still less can formulas of this type serve as the basis for a useful model for text understanding. What makes a text easy or difficult for individual readers is the topic of further research which urgently needs to be done. Because of the highly interactive nature of language understanding, we are confident that it will not prove possible to incorporate the results of this research into procedures of appraising the comprehensibility of texts that look like traditional readability formulas. And we do not think that the

goal of such research should be to produce new formulas. If texts must be changed so that the intended readers can understand them, we want to be able to identify what the barriers are and what improvements actually increase comprehension. If the goal is not to alter the text, we want to be able to convey to the readers how best to approach a text and to deal most efficiently with its complexities.

References

Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), Advances in reading/language research. Greenwich, CT: JAI Press.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), Reading comprehension and education. Newark, DE: International Reading Association.

Anderson, R. C., Mason, J., & Shirey, L. (1984). The reading group: An experimental analysis of a labyrinth. Reading Research Quarterly, 20(1), 6-38.

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. American Educational Research Journal, 14, 367-382.

Anderson, R. C., Shirey, L., Wilson, P., & Fielding, L. (1986). Interestingness of children's reading material. In R. Snow & M. Farr (Eds.), Aptitude learning and instruction (Vol. 4). Hillsdale, NJ: Erlbaum.

Asher, S. R. (1979). Influence of topic interest on black children's and white children's reading comprehension. Child Development, 50, 686-690.

Asher, S. R. (1980). Topic interest and children's reading comprehension. In R. Spiro, B. Bruce, & W. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, NJ: Erlbaum.

Asher, S. R., & Geraci, R. L. (1980). Topic interest, external incentive and reading comprehension. Unpublished manuscript, University of Illinois, Urbana-Champaign.

Asher, S. R., Hymel, S., & Wigfield, A. (1978). Influence of topic interest on children's reading comprehension. Journal of Reading Behavior, 10, 35-47.

Asher, S. R., & Markell, R. A. (1974). Sex differences in comprehension of high- and low-interest reading material. Journal of Educational Psychology, 66, 680-687.

Baldwin, R. S., Peleg-Bruckner, Z., & McClintock, A. H. (1985). Effects of topic interest and prior knowledge on reading comprehension. Reading Research Quarterly, 20, 497-504.

Beck, I., McCaslin, E., & McKeown, M. (1980). The rationale and design of a program to teach vocabulary to fourth-grade students. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Beck, I. L., McKeown, M. G., Omanson, R. C., & Pople, M. T. (1984). Improving the comprehensibility of stories: The effects of revisions that improve coherence. Reading Research Quarterly, 19(3), 263-277.

Beck, I., Perfetti, C., & McKeown, M. (1982). The effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.

Berko, J. (1958). The child's learning of English morphology. Word, 14, 150-177.

Bettleheim, B. (1976). The uses of enchantment: The meaning and importance of fairy tales. New York: Knopf.

Bever, T. G., & Townshend, D. J. (1979). Perceptual mechanisms and formal properties of main and subordinate clauses. In W. E. Cooper & E. C. T. Walker (Eds.), Sentence processing. Hillsdale, NJ: Erlbaum.

Blom, G. E., Waite, R. R., & Zimet, S. G. (1970). A motivational content analysis of children's primers. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books.

Bormuth, J. R. (1966). Readability: A new approach. Reading Research Quarterly, 1, 79-132.

Botel, M., Dawkins, J., & Granowsky, A. (1973). A syntactic complexity formula. In W. H. MacGinitie (Ed.), Assessment of problems in reading. Newark, DE: International Reading Association.

Botel, M., & Granowsky, A. (1972). A formula for measuring syntactic complexity: A directional effort. Elementary English, 49, 513-516.

Bransford, J. D., Barclay, J. R., & Franks, J. (1972). Sentence memory: A constructive versus interpretative approach. Cognitive Psychology, 3, 193-209.

Bransford, J. D., & Franks, J. (1971). The abstraction of linguistic ideas: A review. Cognition, 1, 211-149.

Bruce, B. (1984). A new point of view on children's stories. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), Learning to read in American schools. Hillsdale, NJ: Erlbaum.

Carroll, J. B., Davies, P., & Richman, B. (1971). The American Heritage word frequency book. Boston: Houghton Mifflin.

Cazden, C. (1968). The acquisition of noun and verb inflections. Child Development, 39, 433-448.

Chall, J. S. (1958). Readability: An appraisal of research and appliction. Columbus: Ohio State University, Bureau of Educational Research.

Charrow, R. P. & Charrow, V. R. (1979). Making legal language understandable: A psycholinguistic study of jury instructions. Columbia Law Review, 79(7), 1306-1374.

Davison, A. (1984). Syntactic markedness and the definition of sentence topic. Language, 60(4), 797-846.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. Reading Research Quarterly, 17, 187-209.

Davison, A., & Lutz, R. (1984). Syntactic complexity and discourse context. In D. R. Dowty, L. Karttunen & A. M. Zwicky (Eds.), Natural language processing: Psychological, computational and theoretical perspectives. Cambridge, MA: Cambridge University Press.

Davison, A., Wilson, P., & Herman, G. (ms). Effects of syntactic connectives and organizing cues on text comprehension. Champaign, IL: Center for the Study of Reading.

Dawkins, J. (1975). Syntax and readability. Newark, DE: International Reading Association.

Frazier, L. (1984). Syntactic complexity. In D. R. Dowty, L. Karttunen & A. M. Zwicky (Eds.), Natural language parsing: Psychological, computational and theoretical perspectives. Cambridge MA: Cambridge University Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. Cognition, 6, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14, 178-210.

Freebody, P., & Anderson, R. C. (1983a). Effects on text comprehension of differing proportions and locations of difficult vocabulary. Journal of Reading Behavior, 15(3), 19-39.

Freebody, P., & Anderson, R. C. (1983b). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. Reading Research Quarterly, 18, 277-294.

Glucksberg, S., Trabasso, T., & Wald, J. (1973). Linguistic structures and mental operations. Cognitive Psychology, 5, 338-370.

Graves, M. F., & Slater, W. H. (1986). Could textbooks be better written and would it make a difference? American Educator, 10(1), 36-42.

Gray, W. W., & Leary, B. E. (1935). What makes a book readable? Chicago, IL: University of Chicago Press.

Green, G. M., & Laff, M. O. (1980). Five-year-olds' recognition of authorship by literary style (Tech. Rep. No. 181). Urbana: University of Illinois, Center for the Study of Reading.

Green, G. M., & Olsen, M. S. (1986). Preferences for and comprehension of original and readability-adapted materials. This volume.

Hamburger, H., & Crain, S. (1981). Relative acquisition. In S. Kuczaj (Ed.), Language development: Syntax and semantics. Hillsdale, NJ: Erlbaum.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process of comprehension. Journal of Verbal Learning and Verbal Behavior, 13, 512-521.

Irwin, J. W. (1980). The effects of explicitness and clause order on the comprehension of reversible causal relationships. Reading Research Quarterly, 15, 477-488.

Irwin, J. W., & Pulver, C. J. (1984). Effects of explicitness, clause order and reversibility on children's comprehension of causal relationships. Journal of Educational Psychology, 76(3), 399-407.

Jarvella, R. (1971). Syntactic processing of connected speech. Journal of Verbal Learning and Verbal Behavior, 10, 409-416.

Jenkins, J. R., Pany, D., & Schreck, J. (1978). Vocabulary and reading comprehension: Instructional effects (Tech. Rep. No. 100). Urbana: University of Illinois, Center for the Study of Reading. (ERIC Document Reproduction Service No. ED 160 999).

Jose, P., & Brewer, W. F. (1983). The development of story liking: Character identification, suspense, and outcome resolution (Tech. Rep. No. 291). Urbana: University of Illinois, Center for the Study of Reading.

Kemper, S. (to appear-a). Imitation of complex grammatical constructions by elderly adults. Applied Psycholinguistics.

Kemper, S. (to appear-b). Geriatric psycholinguistics: Syntactic limitations of oral and written language. In L. Light & D. Burke (Eds.), Language and memory in old age. Cambridge, MA: Cambridge University Press.

Klare, G. R. (1963). The measurement of readability. Ames: Iowa State University Press.

Klare, G. R. (1974-75). Assessing readability. Reading Research Quarterly, 10, 62-102.

Klare, G. R. (1984). Readability. In P. D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), Handbook of reading research. New York: Longman.

Kynette, D., & Kemper, S. (to appear). Aging and the loss of grammatical forms. Language and Communication.

Lipson, M. Y. (1983). The influence of religious affiliation on children's memory for text information. Reading Research Quarterly, 18, 448-457.

Marslen-Wilson, W. R. (1975). The limited compatibility of linguistic and functional explanations. In R. Grossman, L. J. San, & T. J. Vance (Eds.), Papers from the Parasession on Functionalism (pp. 409-420). Chicago: University of Chicago, Chicago Linguistic Society.

Marslen-Wilson, W. R., Tyler, L. K., & Seidenberg, M. (1980). Sentence processing and the clause boundary. In W. J. M. Levelt, & G. B. Flores D'Arcais (Eds.), Studies in the perception of language. New York: Wiley.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension. Journal of Reading Behavior, 15, 3-18.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? Reading Research Quarterly, 14, 304-330.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. American Educational Research Journal, in press. (Tech. Rep. No. 347). Urbana: University of Illinois, Center for the Study of Reading.

Olson, D. R., & Filby, N. (1972). On the comprehension of active and passive sentences. Cognitive Psychology, 3, 361-381.

Pearson, P. D. (1974-75). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. Reading Research Quarterly, 10, 155-192.

Pearson, P. D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. Journal of Reading Behavior, 11, 201-210.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Erlbaum.

Reynolds, R. E., Taylor, M. A., Steffensen, M. S., Shirey, L. L., & Anderson, R. C. (1982). Cultural schemata and reading comprehension. Reading Research Quarterly, 17, 353-632.

Rodriguez, N., & Hansen, L. H. (1975). Performance of readability formulas under conditions of restricted ability level and restricted difficulty of materials. Journal of Experimental Education, 44, 8-14.

Sheldon, A. (1974). On strategies for processing relative clauses: A comparison of children and adults. Minnesota Working Papers in Linguistics and Philosophy, No. 2. Linguistic Department, University of Minnesota.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain related information for individuals with high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 18, 275-290.

Spiro, R. J., Bruce, B. C., & Brewer, W. F. (Eds.), (1980). Theoretical issues in reading comprehension. Hillsdale, NJ: Erlbaum.

Steffensen, M. S., Joag-dev, C., & Anderson, R. C. (1979). A cross-cultural perspective on reading comprehension. Reading Research Quarterly, 15, 10-29.

Sticht, T., Armijo, L., Weitzman, R., Koffman, N., Roberson, K., Chang, F., & Moracco, J. (1986). Progress Report. U.S. Naval Post-graduate School.

Tanenhaus, M. K., & Carlson, G. N. (1985). Processing anaphors. Papers from the 1984 NELS meeting. Amherst, MA: Graduate Linguistics Students Association, University of Massachusetts.

Thorndike, E. L. (1921). The teacher's word book. New York: Teacher's College, Columbia University.

Tyler, A., & Nagy, W. (1985). The role of derivational suffixes in sentence comprehension (Tech. Rep. No. 357). Urbana: University of Illinois, Center for the Study of Reading.

Tyler, A., & Nagy, W. (1986). The acquisition of English derivational morphology: Implications for models of the lexicon. Paper presented at the Milwaukee Morphology Meeting/Fifteenth Annual UWM Linguistics Symposium, Milwaukee, WI.

Vogel, M., & Washburne, C. (1928). An objective method of determining grade placement of children's reading material. Elementary School Journal, 28, 373-381.

Yngve, V. H. A. (1960). A model and an hypothesis for language structure. Proceedings of the American Philosophical Society, 104, 444-466.